Thesis/
Reports
Pickens,
J. B.

Use of Chance-Constrained Programming to Account for Stochastic Variation in the A-Matrix of Large-Scale Linear Programs: A Forestry Application

by James B. Pickens

Use of Chance-Constrained Programming to Account for Stochastic Variation in the A-Matrix of Large-Scale Linear Programs: A Forestry Application

A progress report prepared as documentation of completion of Stage 5 for United States Department of Agriculture Cooperative Agreement Number 28-C9-511.

By: James B. Pickens
Associate Professor
School of Forestry and Wood Products
Michigan Technological University
Houghton, MI 49931

August 6, 1992

Use of Chance-Constrained Programming to Account for Stochastic Variation in the A-Matrix of Large-Scale Linear Programs: A Forestry Application

ABSTRACT

Linear programming (LP) is widely used to select the manner in which forest lands are managed. Because of the nature of forestry, this application has several unique characteristics. For example, the models consider many different management actions that take place over many years, thus resulting in very large LP formulations with diverse data. In addition, almost none of the data are known with certainty. The most pervasive occurrence of stochastic information is in the production coefficients, which indicate the uncertain response of the managed forest ecosystem to various management options. A "chance-constrained" approach to handling this uncertainty would often be appropriate in forestry applications--managers and decision makers would like to specify a probability with which uncertain constraints are met. Unfortunately, chance-constrained procedures for A-matrix uncertainty produce nonlinear programming problems, which cannot currently be solved for large-scale forestry applications. This paper utilizes a Monte Carlo simulation approach (a linear program is repeatedly solved with randomly perturbed A-matrix coefficients) to describe the distribution of total output when the individual production coefficients are random. An iterative procedure for "chance-constraining" feasibility is developed and demonstrated with this sort of random A-matrix. An iterative approach is required because the mean and variance of total output are unknown functions of the random A-matrix coefficients and the level of output required. This approach may have applications in other fields as well.

Keywords: Wildland allocation, stochastic production estimates, stochastic A-matrix, chance-constrained programming.

Use of Chance-Constrained Programming to Account for Stochastic Variation in the A-Matrix of Large-Scale Linear Programs: A Forestry Application

0. THE FORESTRY OPTIMAL LAND ALLOCATION PROBLEM

Operations research methods have been employed to solve a wide range of forestry problems. Many of these applications, such as manufacturing processes and transportation problems, are very similar to applications in other fields. However, one important application in forestry results in linear programming formulations that are rather distinctive. This application deals with the selection of optimal land management (or "land allocation") for a forest (Ware and Clutter [13], Johnson and Scheurman [4]).

There are several factors which distinguish the land allocation problem from other operations research applications. First, the land resource is typically divided into a necessarily large number of homogeneous units (which will be referred to as analysis areas) based on biological productivity, age of the current timber stand, current timber volume, species composition, and many other factors. Thus, these models have a large number of constraints to limit the amount of land allocated to that available by analysis area. In addition, many constraints are often necessary to restrict the way in which land can be managed. The total number of constraints in practical applications ranges from 300 to 10,000 or more, and most models have several thousand.

Second, a wide variety of alternative management options are typically available. Thus, the land allocation models have a large number of choice variables as well. Ware and Clutter [13] present an example where only timber

is considered and only one management system (clearcut harvests) is available, but there are still 54 timing choices. Consideration of multiple-use management, alternatives other than clearcutting for timber, and accounting for multiple analysis areas greatly inflate this number. The total number of choice variables in practical applications ranges from 1000 to 100,000, and most models have tens of thousands.

Third, nearly all of the coefficients of the land allocation linear program are not known with certainty. The uncertainty occurs in the objective function and the right-hand-side (RHS) vector; however, it is A-matrix uncertainty that is the most common and the most troublesome to handle in an optimization context. The size of the forest land allocation models precludes nonlinear programming approaches such as those developed in Miller and Wagner [5] and Jagannathan [3]. A-matrix uncertainty results from estimation errors as well as the intrinsically uncertain response of the natural resource system to management actions. For example, current timber volume available for harvest is typically estimated using very low-intensity sampling methods. Future volume is estimated by projection of the estimated current volume using statistically derived models ("growth and yield equations"), and is never known with certainty. Furthermore, the uncertain production information is generally never known with certainty before the relevant decisions are made and management actions implemented. For example, the decision to harvest a stand of timber is made before an accurate pre-sale survey is conducted to generate a precise estimate of the volume. Thus, two-stage methods which rely on better information that becomes available later are not appropriate (e.g. -- see Wagner [12]).

One approach to this type of randomness is to passively study the characteristics of the model's solution in the presence of specific types of

uncertainty (Sengupta [9]). Pickens and Dress [7] used this approach to study the case of stochastic production estimates in a land allocation application, and characterized the distributions of the objective function value reported by the linear program and the actual objective function value when the solution is implemented. This study also revealed that the chosen optimal solution is very unlikely to satisfy constraints with many stochastic coefficients. Ninety percent of the optimal solutions observed with stochastic production coefficients in only one constraint were not actually feasible in terms of meeting the desired production level. In addition, the interaction between feasibility and optimality was demonstrated by the fact that 6 of the 30 simulations resulted in actual objective function values higher than the objective function value in the presence of perfect information. Clearly, all 6 of these solutions were infeasible.

This paper presents a simulation approach to optimizing forestry land allocations models with random A-matrix coefficients. In section 1, three "levels" of models and a number of variables associated with those levels are defined. Section 2 discusses a case study introduced to demonstrate the simulation approach. Two possible cases are identified and addressed. Section 2.1 discusses the first case, which occurs when the output enters the formulation only through the objective function, and thus feasibility is assured in the presence of stochastic production estimates. In this case, uncertainty causes timber production to be optimistically estimated by the linear program, and a method to predict the distribution of the actual level of production to be expected from a given solution is developed. Section 2.2 presents the second case, where the output enters both through the objective function and as a minimum production constraint. This case is most common in forestry land allocation problems. In this case the feasibility of the linear program's optimal solution is not assured, and an iterative method to

chance-constrain feasibility is developed. Section 3 discusses the generality of the study's results.

1. DEFINITIONS

The paper utilizes three "levels" of models. The models are:

1. The level 1 model is reality. In this model, production now and in the future is assumed known with certainty.

2. The level 2 model is generated from the level 1 model by adding an independent stochastic component to each level 1 timber harvest coefficient.

3. The level 3 model is generated from a given level 2 model by adding an independent stochastic component to each level 2 timber harvest coefficient.

The variables of interest are defined as Y_{ij}, and represent the calculated harvest when the i level A-matrix and the j level optimal solution (decision variable activities) are combined. If i and j are not the same (from the same optimal solution), the model with the higher index was generated from the model with the lower index. For example, a Y_{23} observation uses a level 3 model generated from the level 2 model whose A-matrix the level 3 solution is imposed upon.

The specific variables of interest are Y_{12}, Y_{22}, Y_{23}, and Y_{33}. Y_{12} represents the total harvest calculated when the level 2 optimal solution is imposed on the level 1 A-matrix. Thus, Y_{12} is the actual harvest that will occur when the level 2 solution is implemented. Y_{22} is the harvest in the

linear program optimal solution report for the level 2 model. Thus, Y_{22} is the amount of harvest the practitioner would expect based on the linear program solution report and Y_{12} is the amount that would actually be harvested. Y_{23} and Y_{33} have a similar relationship, but are one level up in the hierarchical modelling system. In real-world applications the level 1 model, which has correct production coefficients, is never known, and only one level 2 model is available. Therefore, we wish to infer the relationship between Y_{12} and Y_{22} based on simulations with only one level 2 model, from which any number of level 3 models can be simulated.

Let the notation Y_{ii} indicate that i and j are equal (i.e. -- Y_{22} and Y_{33}), while the notation $Y_{(i-1)i}$ indicates that the value of the two subscripts are one level separated (i.e. -- Y_{12} and Y_{23}). The subscript i for the variables Y_{ii} and $Y_{(i-1)i}$ indicates the model level for the optimal solution (decision variable activities) used to calculate the variable, and will be used for the remainder of the paper.

One question central to the problem is the nature of the relationship between Y_{ii} and $Y_{(i-1)i}$. With stochastic production coefficients, the decision variables are also random variables (Sengupta [10]). In addition, because the activity level of the decision variables is selected by an optimization process, the linear programming solution process will systematically seek optimistic estimates and avoid pessimistic estimates. Therefore, the estimate Y_{ii} will tend to be an optimistically biased estimate of $Y_{(i-1)i}$, leading a practitioner solving linear programs with stochastic A-matrix estimates to anticipate higher production than the solution will produce (in expectation). That is, Y_{22} is a biased estimate of Y_{12}. Figure 1 shows this relationship for 30 simulations of the level 2 case study model discussed below.

For convenience, we define the following variables:

$$B_i = Y_{ii} - Y_{(i-1)i} \qquad i=2,3 \qquad (1)$$

$$\%B_i = ((Y_{ii} - Y_{(i-1)i}) / Y_{ii})\ 100 = (B_i / Y_{ii})\ 100 \qquad i=2,3 \qquad (2)$$

B_i is the bias associated with using Y_{ii} to estimate $Y_{(i-1)i}$, and $\%B_i$ is the proportionate bias for that estimation. The subscript i for the variables B_i and $\%B_i$ indicates the model level for the optimal solution used to calculate the variable.

The distributions of the variables Y_{12}, Y_{22}, Y_{23}, Y_{33}, B_i, and $\%B_i$ are not, in general, known. However, in the forestry land allocation problem, the extremely large size of the models and the small contribution of any one analysis area (and therefore any decision variable activity) suggests that these variables might tend toward normality. The tendency toward normality is evident from the shape of the frequency distribution for both Y_{12} and Y_{22} in Figure 1. The hypothesis of normality is tested in the case study below.

2. A CASE STUDY

A case study model was formulated using data collected by the USDA Forest Service [11]. This data includes production, cost, and land base information for all renewable resource lands in the United States. Only a portion of the data was used in this study.

Two outputs, timber harvest and domestic grazing, were included in the model. The data coefficients in the model are the average annual production of wood harvest in cubic feet and the average annual production of grazing by domestic animals measured in a standardized unit, the animal unit month, both

over a time horizon of the next 50 years. The timber output is assumed to be random, while the domestic grazing is assumed known with certainty.

It is often the case in biological systems that the variability of a parameter is proportional to the magnitude of the parameter. Furthermore, because many factors affect the development of a stand of timber, and, excluding catastrophes such as fire, none of these factors dominate the resulting production, it is not unreasonable to expect normally distributed errors in timber yield estimation. The assumed error distribution for the simulations was thus normal with a coefficient of variation of 0.10. This assumption is critical because the simulation approach discussed below requires that the process which creates the production errors used to simulate the level 3 models from the level 2 models must be, in some sense, the same as the process which generated the level 2 models from the level 1 model. This error structure insures that this requirement is met because a proportioning relationship would be expected between Y_{12} and Y_{22}, and also between Y_{23} and Y_{33}. As tested below, the variable $\%B_i$ (Equation 2) will be used to infer the distribution of actual production from the level 2 solution and the level 3 simulations. If the A-matrix error structure defined above is not acceptable, an alternative measure must be developed which is suitable for the new error structure.

The portion of the Nation's renewable resource land base selected for inclusion in the case study model was all lands in Oregon except those owned by the Forest Service. 292 analysis areas are defined, based on the classification system described in USDA Forest Service [11]. Characteristics which defined separate analysis areas were ecological community, age, biological productivity, and ownership category. This number (292) of analysis areas was considered large enough to produce reasonable large model

asymptotic results, and small enough to allow the large number of simulations needed. About 10 management alternatives, which are represented in the model as decision variables, are available for each analysis area. These options represent differing emphases on timber, domestic grazing, and wildlife products.

The original data was assumed to be correct, and thus served as the level 1 model in the model testing framework. 374 million cubic feet of wood were selected annually for harvest when this model was optimized using the maximize present net value objective function, while the land base was able to produce a maximum timber harvest of 505 million cubic feet.

Next, 30 level 2 models were generated from the level 1 model using the International Mathematical and Statistical Library (IMSL) of subroutines for random number generation [2]. The subroutines were used to generate standardized normal random variables, which were then transformed to produce a coefficient of variation of 0.10. The system clock was accessed to set the seed for random number generation. Finally, six level 2 models were selected at random, and 30 level 3 models were generated for each of the chosen level 2 models.

The algebraic form of the linear program for each of the three levels is:

$$\text{Maximize } z = \sum_{k=1}^{2} v_k t_k - \sum_{i=1}^{292} \sum_{j=1}^{n_i} c_{ij} x_{ij} \qquad (3)$$

Subject to:

$$\sum_{i=1}^{292} \sum_{j=1}^{n_i} p_{ijk} x_{ij} - t_k = 0 \qquad k = 1, 2 \qquad (4)$$

$$t_1 \geq b_1 \qquad (5)$$

$$\sum_{j=1}^{n_i} x_{ij} = 1_i \qquad i = 1, 2, \ldots, 292 \qquad (6)$$

where:

b_1 = the timber harvest requirement

c_{ij} = the cost of treating 1 acre of analysis area i using management option j

l_i = the number of acres of analysis area i

n_i = the number of management options for analysis area i

p_{ijk} = the production of product k (timber harvest = 1 and domestic grazing = 2) by analysis area i when treated using management option j

t_k = an accounting variable representing the total production of product k

v_k = the value of one unit of product k

x_{ij} = the number of acres of analysis area i treated with management option j. The x_{ij} are the decision variables.

Equation 3 represents the present net value associated with any set of the decision variables. The first summation represents the revenue, while the second summation represents costs. Equations 4 accumulate the total production for each of the two products and create an accounting variable (t_k) whose value is the total production of the k^{th} product. Equation 5 limits the timber harvest to be at least as large as b_1. Equations 6 are land area constraints, and limit the allocated land to the available amounts in each analysis area. The random components enter the level 2 and 3 models as the p_{ij1} in Equation 4.

Two distinct cases are of interest to land managers. In the first case no desired level of timber harvest is specified ($b_1 = 0$), while in the second case a minimum amount of harvested volume is required ($b_i > 0$). Case 1 involves a single set of simulations to characterize the distribution of Y_{12}, while case 2 requires an iterative process. The first step of the iterative

process characterizes the distribution of Y_{12}, and the second step checks whether the chance-constraint is satisfied. If the chance-constraint is not satisfied, b_1 is adjusted and steps one and two are repeated. The two cases are addressed in sections 2.1 and 2.2 respectively.

2.1 Case 1: Optimization without a Minimum Timber Harvest Constraint

Although it is far more common that land allocation models include timber harvest both in the objective function and constraints, it is sometimes the case that timber harvest is represented only in the objective function. In this case it is usually quite important to know the actual timber harvest that will occur when the level 2 solution is implemented (Y_{12}). The land management organization might, using the optimistically biased harvest reported by the level 2 model, commit to sell more timber than will be harvested or expand processing facilities to accommodate the Y_{22} volume. In either of these cases, serious penalties could result from the use of the optimistically biased Y_{22} reported by the linear program to estimate Y_{12}. Thus, it is important to predict Y_{12} even if *a priori* timber harvest constraints are not specified.

The thirty level 2 models and the six sets of 30 level 3 models generated from randomly selected level 2 models were optimized without a timber harvest constraint ($b_1 = 0$). Three parameters, Y_{ii}, $Y_{(i-1)i}$, and $\%B_i$ (Equation 2) were observed for each solution. The Statistical Analysis System (SAS [8]) was used for all statistical tests, and the Functional Mathematical Programming System (FMPS) optimization software available on UNISYS computers was used to optimize the linear programs.

2.1.1 Normality tests

The argument was made above that many of the parameters from these large models would tend to be normally distributed. Table 1 presents the results of normality tests for the observed variables using the Shapiro-Wilks statistic. Each tabulated value is the probability of observing the simulated sample given the null hypothesis that this sample was derived from a normal population. Thus, low values indicate that it is unlikely that the data were produced by a process that generates normal random variables. The seven columns in the table represent the level 2 simulation and the six level 3 simulations generated from randomly selected level 2 models.

Using a type I error level of 5 percent (i.e. -- $\alpha = .05$), two of the 21 tests are significant. Although this is somewhat above the expected number of hypothesis rejections of 1.05, it does indicate that these simulations are generating random variables that at least tend strongly toward normality. The slight departure from normality, if in fact there is any, could be explained by the size of our test case model (292 analysis areas), and might not be expected for typical land allocation models with even more analysis areas.

Table 2 presents the simulation means in a format identical to Table 1. The units for $Y_{(i-1)i}$ and Y_{ii} are million cubic feet per year of harvested timber, while the units of the reported values for $\%B_i$ are percent. The percent bias values reported are all significantly different from zero ($p = .001$). Student's t statistics to test whether the percent bias is equal to zero ranged from 4.03 to 6.92. This confirms that using Y_{ii} to estimate $Y_{(i-1)i}$ will result in optimistically biased estimates.

2.1.2 Predicting Y_{12}

Next, we wish to find a method to estimate the Y_{12} from one Y_{22} and its associated level 3 simulations. In practice, only one sample from a population of possible level 2 models is available. Thus, a function relating the one Y_{22} to the simulated Y_{23} and Y_{33} is needed that is not sensitive to the level 2 model selected. $Y_{(i-1)i}$, Y_{ii}, and $\%B_i$ were analyzed using a variance components (also called random effects or Model II) analysis of variance (ANOVA) with a completely randomized design to determine if the i-1 level model was an important predictor of each variable at the i^{th} level. For details on random effects models, see Montgomery [6] or other introductory experimental design text. The treatments in the analysis were the level 2 model (which was generated from the level 1 model) and the six level 3 models (each simulated from a different level 2 model). Each treatment had 30 observations, yielding an analysis with a total of 210 observations. The statistical model is:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (7)$$

where y_{ij} is the dependent variable, τ_i is the treatment effect, and ε_{ij} is the random error associated with each observation.

The two analyses using $Y_{(i-1)i}$ and Y_{ii} as dependent variables indicated that the treatments were highly significant ($p = .0001$) predictors. This result, that the model which was used to generated a higher level model is significant to the total production estimates, is certainly not surprising.

However, the ANOVA using $\%B_i$ as the dependent variable could determine no difference between the same seven treatments ($p=.2839$). This indicates that proportionate bias observed at any level will not vary when errors are simulated using coefficient of variation in the unconstrained models. This is a very useful result, and leads to the following equation to predict the

actual production in the presence of stochastic production coefficients not involved in constraints:

$$Y_{12} = Y_{22} \left(1 - \%B_3 / 100 \right) \quad (8)$$

Equation 8 can be rearranged to yield:

$$Y_{12} / Y_{22} = (1 - \%B_3 / 100) \quad (9)$$

The $\%B_3$ was estimated using the 180 level 3 simulations to have a mean of 4.35 and a standard deviation of 4.11. In addition, the Kolmogorov-Smirnov test could not detect divergence from normality ($p = .145$). Therefore, the right hand side of Equation 9 appears to be distributed normally with estimated mean 0.9565 and estimated standard deviation of 0.0411. In addition, the parameter estimates for the distribution can be improved by increasing the number of level 3 simulations.

2.2 Case 2: Optimization with a Minimum Timber Harvest Constraint

It is common for forestry land allocation models with the maximize present net value objective to include timber value in the objective function and to constrain the timber harvest required above a certain amount. The reasons for imposing these constraints vary, but generally reflect a belief that the forest is responsible for supporting an external need. The responsibility is usually to supply fiber to a mill when the forest is owned by a forest products company, while publicly owned forests are often required to provide a secure source of raw materials to support community stability, provide employment, and secure (in some sense) the national welfare.

Feasibility of the optimal solution is not assured when timber harvest constraints are imposed in linear programs with stochastic timber production

coefficients. The problem addressed in this section is to determine the value b_1 (Equation 5) to "chance-constrain" feasibility of the model at a pre-specified level. The classic presentation of chance-constraints by Charnes and Cooper (e.g. -- [1]) deals with random right-hand-sides (among other topics). In our case, we wish to adjust the RHS (b_1) not in response to uncertainty about b_1, but in response to uncertainty in the A-matrix of large-scale linear programs. In this example we assume that the decision maker wishes to be 80 percent sure that at least 450 million cubic feet are harvested. Specifically, we wish to determine b_1 in the level 2 problem such that:

$$P(Y_{12} \geq 450{,}000{,}000) = .80 \qquad (10)$$

Three results are required to chance-constrain this problem. First, we must confirm that certain random variables ($Y_{(i-1)i}$ and $\%B_i$) generated by the constrained simulations appear normally distributed. Second, we must confirm that these variables are not dependent on the i-1 level model which was used to simulate i level problems. Third, an iterative approach is required to solve this problem because the relationship between b_1 and the distribution of Y_{12} can only be estimated, and may change with different b_1 values. Thus, iterations involving alterations to b_1 and estimation of the Y_{12} distribution must continue until the estimated Y_{12} distribution at the k^{th} iteration is "sufficiently close" to the estimated distribution at the k-1 iteration.

If the timber harvest constraint is always binding in the simulations, Y_{ii} is a constant. Then, by Equation 2, $\%B_i$ is a linear deterministic transformation of $Y_{(i-1)i}$, and $\%B_i$ and $Y_{(i-1)i}$ can be used interchangeably to represent the variability of the system. This result is not a requirement of

the general procedure outlined above, but will be used in this example to simplify the discussion.

2.2.1 Iteration one

To chance-constrain the case study model, a starting point was chosen which ignores two facts; Y_{ii} is a biased estimate of $Y_{(i-1)i}$, and $Y_{(i-1)i}$ is a random variable. The 30 level 2 simulations and the six sets of 30 level 3 simulations were optimized using an initial RHS of the desired production (450,000,000).

2.2.1.1 Normality tests

The resulting values for $Y_{(i-1)i}$ and $\%B_i$ were evaluated for the seven sets of simulations, as in the unconstrained case, to determine if they appear to be normal random variables. The variable Y_{ii} was not tested because all values which were previously below 450 million cubic feet are now forced to be 450 million, resulting in a truncation of the original distribution. Only 2 of the 210 simulations produced slack harvest for the timber constraint, so normality can no longer be expected.

Table 3 presents the results of normality tests (analogous to Table 1) for the observed variables using the Shapiro-Wilks statistic. The probability associated with the Shapiro-Wilks tests were identical for $Y_{(i-1)i}$ and $\%B_i$ for five of the seven simulations. This occurred because Y_{ii} was constant (450 million) for all 30 simulations. The two simulations where these probabilities differed each had one replicate where Y_{ii} was greater than the required 450 million cubic feet of harvest. One of the five pairs of identical probabilities was significant at $\alpha = .05$, while none of the four separate tests were significant. Again, the number of rejected hypotheses is

slightly greater than would be expected. However, none of the insignificant tests even approached the .05 level required to reject the null hypothesis of normality. It seems likely that the process being modelled generates normal random deviates.

2.2.1.2 ANOVA tests

Table 4 is analogous to Table 2 in the unconstrained case, and presents the mean values for $Y_{(i-1)i}$ and $\%B_i$. Model 2 ANOVA analyses were conducted for the dependent variables $Y_{(i-1)i}$ and $\%B_i$ to determine if the seven treatments (the level 2 simulation and the six selected level 3 simulations) were significant predictors. Equation 7 presents the relevant linear statistical model for this case as well. The treatments did not contribute significantly for either dependent variable ($p=.3012$ for $Y_{(i-1)i}$ and $p=.2814$ for $\%B_i$). This indicates that the level 3 simulations from any level 2 model can be used to make inferences concerning the $\%B_2$ of the level 2 solution when a minimum timber harvest is required.

2.2.2 Iteration two

Next, the timber harvest constraint level (b_1) for the level 3 simulations from the first randomly selected level 2 model was modified to compensate for the bias and uncertainty of the system. The RHS was modified to reflect:

1. The bias caused by using Y_{ii} to estimate $Y_{(i-1)i}$.
2. The desire to be 80 percent sure that Y_{12} is at least 450 million cubic feet.

To adjust for (1) above, the $\%B_3$ for the simulations (3.45) was converted to decimal form and multiplied by b_1, then added to b_1 to yield b_1^* (465.5 million). If b_1^* were used to replace b_1, then the bias would be eliminated, but the distribution of Y_{12} would be centered at 450 million. This would imply that the probability of harvesting at least 450 million cubic feet is 0.5.

To adjust for (2) above, the appropriate Z value from the standard normal table to indicate 80 percent assurance (Z = 0.842) was multiplied by the standard deviation of $\%B_3$ (converted to decimal form: s = 0.02973) for the appropriate level 3 simulation, which was then multiplied by the original b_1 and added to b_1^*. The new b_1 value was 476.8 million cubic feet of harvest. A solution to the level 2 model with this timber target would be assured of meeting the chance-constraint except that the distribution of $\%B_3$ with the modified RHS value may be different from the distribution in iteration 1.

Thus, a set of simulations based on the new b_1 was used to test the sensitivity of the distribution of the $\%B_3$ (and therefore Y_{23}) to the modifications in b_1. The level 3 simulations were optimized, and the resulting mean and standard deviation for $\%B_i$ were calculated. The parameter estimates were 3.47 for the mean and 2.81 for the standard deviation. The values observed using the initial estimate of b_1 (450 million) were a mean of 3.45 and a standard deviation of 2.97. Neither the estimated variance nor mean for $\%B_i$ with the revised RHS (b_1 = 476.8 million) were significantly different from the estimates with the initial RHS (b_1 = 450 million). In fact, the calculated statistics (F = 1.17, t = 0.03) were far from the values needed for hypothesis rejection at α = .05. In addition, the difference between the means of $\%B_3$ (3.47 - 3.45 = .02 percent or 0.09 million cubic feet of harvest) was very small when compared with the underlying variability of

the system. Therefore, $\%B_3$ parameters for the k^{th} and $k\text{-}1^{th}$ iterations are sufficiently close to conclude that the chance-constraint is approximately met with b_1=476.8 million cubic feet. In other words, the solution to the level 2 model with b_1 set at 476.8 million cubic feet is the optimal solution that insures an (approximately) 80 percent chance of producing 450 million cubic feet of wood harvest. If the distribution of the new $\%B_3$ had been appreciably different than the previous one, further iterations would have been required.

2.2.3 Evaluation using all 30 level 2 models

One final test of the method involved estimating a new b_1 for each of the 30 level 2 models, then solving the level 2 models and observing the actual Y_{12} for the optimal solutions. The six level 2 models which had been selected to simulate level 3 models employ the results of their level 3 simulations to adjust b_1. The 24 level 2 simulations which were not chosen to generate level 3 simulations used a pooled estimate of all 180 level 3 solutions to calculate the new b_1 level. The calculated b_1 values are presented in Table 5.

When the six level 2 models that were used to generate the level 3 simulations, and therefore had b_1 estimates based on their level 3 simulations, were optimized, five of the six solutions generated the desired chance-constrained harvest of 450 million or more cubic feet. Thus, one of six solutions failed to reach the desired level of harvest. This result fits very well the probability statement in Equation 10.

When the 24 level 2 models that did not serve as the origin of the level 3 simulations were optimized using the pooled estimate of b_1, 20 solutions resulted in Y_{12} values greater than the required 450 million cubic feet, while 3 replications produced less than 450 million cubic feet. One of the level 2 models was infeasible. Thus three of 23, or 13 percent, of the level 2

solutions failed to reach the desired production. This result is somewhat less than the 20 percent failure rate specified by Equation 10. However, it is well within the expected range of likely results for a binomial experiment with 23 replicates and probability of success at 0.8. In fact, the probability of observing more than 20 successes given the null hypothesis that p=.8 is 0.1331, while the probability of observing exactly 20 successes is 0.1633. Thus, the probability of observing 20 or more successes given the probability statement in Equation 10 is .2964. Furthermore, inspection of the 20 simulations that resulted in Y_{12} greater than 450 million reveal several simulations which attained the desired harvest by only a small amount, 100,000 cubic feet in one case and 1.3 million cubic feet in another. The three simulations that failed to produce Y_{12} values of 450 million fell short by 18, 20, and 30 million cubic feet.

3. CONCLUDING REMARKS

A case study model was used to demonstrate solution characteristics when widespread uncertainty occurs in the A-matrix of large-scale linear programs. Although the formulation and data characteristics are specific to forestry applications, it seems likely that the analysis has applications elsewhere. There were three fundamental conclusions reached in the test case.

The first result, which applies to all cases where a random A-matrix is used, merely confirms intuition. Solutions to linear programs with a stochastic A-matrix report optimistically biased results. The bias results from the linear program systematically selecting decision variables with optimistic estimates and avoiding decision variables with pessimistic

estimates. The bias can result in optimistic objective function values, selection of optimal solutions that are truly infeasible, or both.

The second result of the study is that many of the parameters observed when stochastic production coefficients are used in large models will tend toward normality. This conclusion relies on asymptotic statistical results, and therefore is only true for large models. The result also requires that each decision variable contributes only slightly to the total production. In this situation, the total production is the sum of a large number of small independent errors, none of which significantly influence the total, and the random variables generated would approximate a normal distribution. The distributions of Y_{ii} (i=2,3), $Y_{(i-1)i}$ (i=2,3), and $\%B_i$ (i=2,3) were tested using the Shapiro-Wilks test when sample size was less than 50 and the Kolmogorov-Smirnov test for large samples. Most of these tests could find no divergence from normality, and the number of hypothesis rejections observed was within the range to be expected if the data were observations from a normal population.

The third result is that, in the case where errors are appropriately generated by coefficient of variation, the proportionate bias ($\%B_i$) is not affected by the specific model (the one level 1 model or any level 2 model) used to generate the simulations. This result is general for models where no constraints involve the stochastic response coefficients. The result was still true in the case study model when constraints were imposed on the product with stochastic estimates. However, it is not clear that this result will be true for all models which involve stochastic estimates in constraints, and should be tested in future work.

REFERENCES

[1] A. Charnes and W.W. Cooper, Deterministic equivalents for optimizing and satisficing under chance constraints, Operations Research 11 (1963) 18-39.

[2] IMSL, Library Reference Manual (4 volumes),(IMSL Customer Relations, Houston, Texas, 1982).

[3] R. Jagannathan, Chance-constrained programming with joint constraints, Operations Research 22 (1974) 358-372.

[4] K.N. Johnson and H.L. Scheurman, Techniques for prescribing optimal timber harvest and investment under different objectives - discussion and synthesis, Forest Science Monograph 18 (1977) 1-31.

[5] B.L. Miller and H.M. Wagner, Chance constrained programming with joint constraints, Operations Research 13 (1965) 930-945.

[6] D.C. Montgomery, Design and Analysis of Experiments (Second edition, John Wiley and Sons, 1984).

[7] J.B. Pickens and P.E. Dress, Use of stochastic production coefficients in linear programming models: Objective function distribution, feasibility, and dual activities, Forest Science 34 (1988) 574-591.

[8] SAS, SAS User's Guide: Basics (and Statistics), (Version 5, SAS Institute, Cary, NC, 1985).

[9] J.A. Sengupta, Stochastic linear programming with chance constraints, International Economic Review 11 (1970) 101-116.

[10] J.A. Sengupta, Stochastic Programming: Methods and Applications (American Elsevier Publishing Company, New York, 1972).

[11] USDA Forest Service, RPA-79 assessment - book of procedures, framework for supply analysis (USDA Forest Service, Washington, DC, 1977).

[12] H.M. Wagner, Principles of Operations Research with Applications to Managerial Decisions (Prentice-Hall, Englewood Cliffs, New Jersey, 1975).

[13] G.O. Ware and J.L. Clutter, A mathematical programming system for management of industrial forests, Forest Science 17 (1971) 428-445.

Table 1. Distributional results from the model testing system with no timber harvest constraint. The tabulated values are the probability of observing the reported Shapiro-Wilks (SAS [8]) statistic given the null hypothesis that the observations are normally distributed.

	Simulation						
	Level 2	Level 3					
Variable		1	2	3	4	5	6
$Y_{(i-1)i}$	.300	.100	.024*	.549	.402	.824	.481
Y_{ii}	.599	$\leq$.01**	.509	.936	.771	.946	.086
$\%B_i$	.715	.606	.479	.763	.606	.067	.099

*Significantly different from normal at $\alpha = .05$.

**Significantly different from normal at $\alpha = .01$.

Table 2. Observed means for the model testing system with no constraints.

		Simulation						
		Level 2	Level 3					
Variable	Units		1	2	3	4	5	6
$Y_{(i-1)i}$	10^6 ft^3	363	348	352	381	367	357	378
Y_{ii}	10^6 ft^3	378	364	366	397	390	371	400
$\%B_i$	percent	3.70	4.08	3.77	3.85	5.79	3.47	5.13

Table 3. Distributional results from the model testing system with 450 million as the RHS of the timber harvest constraint. The tabulated values are the probability of observing the reported Shapiro-Wilks (SAS [8]) statistic given the null hypothesis that the observations are normally distributed.

	Simulation					
	Level 2	Level 3				
Variable		1	2	3	4	5	6
$Y_{(i-1)i}$	.635	.625	1.000	.897	.545	.019*	.438
$\%B_i$	.635	.625	1.000	.913	.545	.019*	.237

*Significantly different from normal at $\alpha = .05$.

Table 4. Observed means for the model testing system with timber harvest constrained to greater than or equal to 450 million cubic feet.

		Simulation						
		Level 2	Level 3					
Variable	Units		1	2	3	4	5	6
$Y_{(i-1)i}$	10^6 ft^3	433	434	431	434	426	434	427
$\%B_i$	percent	3.78	3.45	4.21	3.63	5.33	3.49	5.18

Table 5. RHS values calculated for timber harvest to chance-constrain the level 2 models. The six level 2 models that were selected for generation of level 3 models use their level 3 simulations to estimate b_1. The 24 level 2 models that were not selected for generation of level 3 models use a pooled estimate of the 180 level 3 simulations to estimate b_1.

Level 2 Model	Calculated b_1 in Million Cubic Feet
1	476.8
2	480.7
3	478.2
4	491.4
5	479.2
6	491.6
Pooled Estimate	483.4

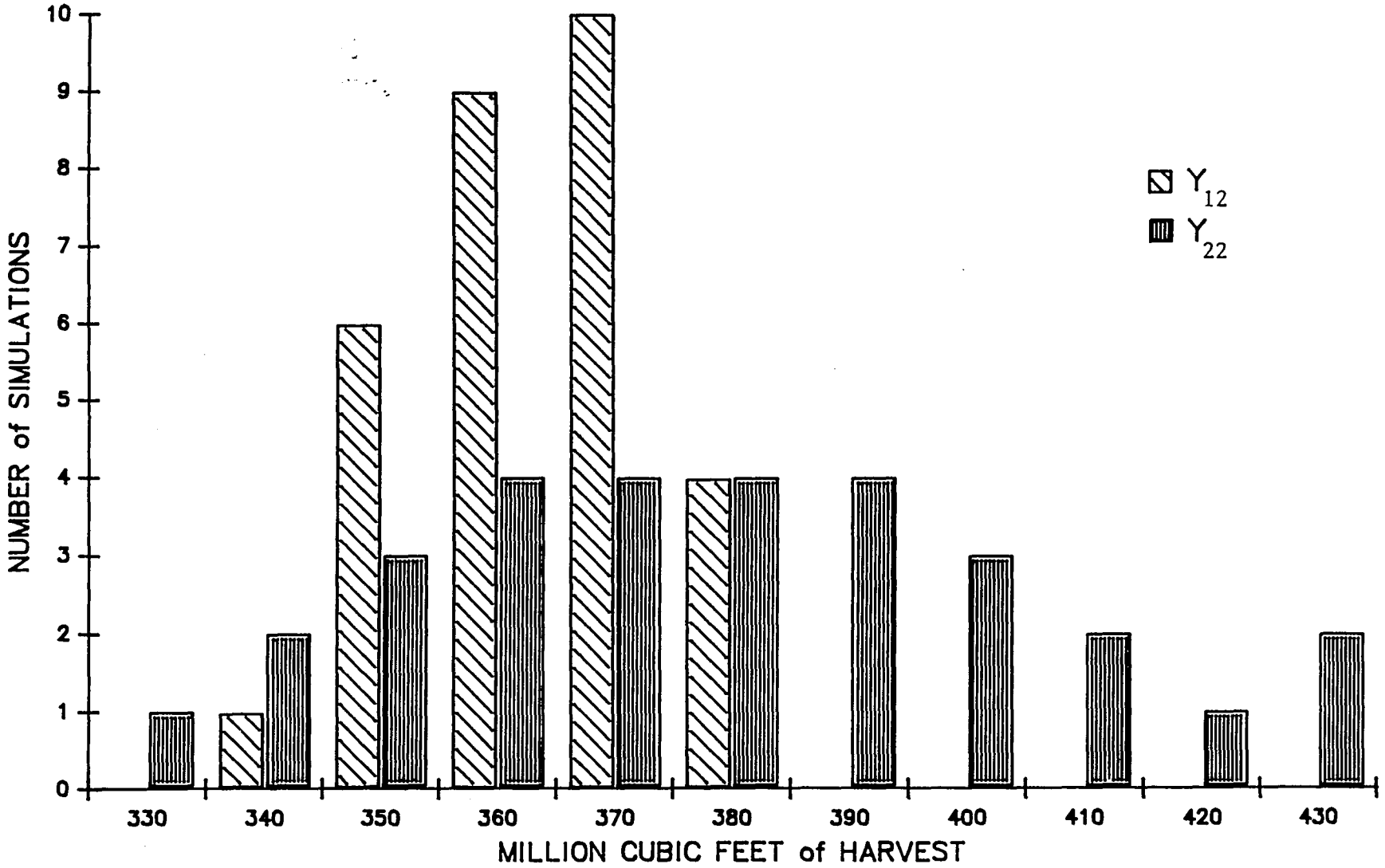

Figure 1. Observed frequencies of Y_{12} and Y_{22} for the 30 level 2 simulations with no timber harvest constraint.